Genotype-specific SNP map based on whole chromosome 3B sequence information from wheat cultivars Arina and Forno

Shatalina, Margarita ; Wicker, Thomas ; Buchmann, Jan P ; Oberhaensli, Simone ; Simková, Hana ; Doležel, Jaroslav ; Keller, Beat

Abstract: Agronomically important traits are frequently controlled by rare, genotype-specific alleles. Such genes can only be mapped in a population derived from the donor genotype. This requires the development of a specific genetic map, which is difficult in wheat because of the low level of polymorphism among elite cultivars. The absence of sufficient polymorphism, the complexity of the hexaploid wheat genome as well as the lack of complete sequence information make the construction of genetic maps with a high density of reproducible and polymorphic markers challenging. We developed a genotype-specific genetic map of chromosome 3B from winter wheat cultivars Arina and Forno. Chromosome 3B was isolated from the two cultivars and then sequenced to 10-fold coverage. This resulted in a single-nucleotide polymorphisms (SNP) database of the complete chromosome. Based on proposed synteny with the Brachypodium model genome and gene annotation, sequences close to coding regions were used for the development of 70 SNP-based markers. They were mapped on a Arina × Forno Recombinant Inbred Lines population and found to be spread over the complete chromosome 3B. While overall synteny was well maintained, numerous exceptions and inversions of syntenic gene order were identified. Additionally, we found that the majority of recombination events occurred in distal parts of chromosome 3B, particularly in hot-spot regions. Compared with the earlier map based on SSR and RFLP markers, the number of markers increased fourfold. The approach presented here allows fast development of genotype-specific polymorphic markers that can be used for mapping and marker-assisted selection.

DOI: https://doi.org/10.1111/pbi.12003

Plant Biotechnology Journal

Plant Biotechnology Journal (2012), pp. 1–10 doi: 10.1111/pbi.12003

Genotype-specific SNP map based on whole chromosome 3B sequence information from wheat cultivars Arina and Forno

Margarita Shatalina[1], Thomas Wicker[1], Jan P. Buchmann[1], Simone Oberhaensli[1], Hana Šimková[2], Jaroslav Doležel[2] and Beat Keller[1,*]

[1]*Institute of Plant Biology, University of Zurich, Zurich, Switzerland*

[2]*Centre of the Region Haná, Institute of Experimental Botany, Olomouc, Czech Republic*

Received 1 June 2012;
revised 27 August 2012;
accepted 30 August 2012.
**Correspondence* (Tel + 41 44 634 8230; fax + 41 44 634 8204; email bkeller@botinst.uzh.ch)
Accession numbers for EMBL seq database: HE996228-HE996964.

Keywords: wheat, genetic mapping, single-nucleotide polymorphism, synteny, Brachypodium.

Summary

Agronomically important traits are frequently controlled by rare, genotype-specific alleles. Such genes can only be mapped in a population derived from the donor genotype. This requires the development of a specific genetic map, which is difficult in wheat because of the low level of polymorphism among elite cultivars. The absence of sufficient polymorphism, the complexity of the hexaploid wheat genome as well as the lack of complete sequence information make the construction of genetic maps with a high density of reproducible and polymorphic markers challenging. We developed a genotype-specific genetic map of chromosome 3B from winter wheat cultivars Arina and Forno. Chromosome 3B was isolated from the two cultivars and then sequenced to 10-fold coverage. This resulted in a single-nucleotide polymorphisms (SNP) database of the complete chromosome. Based on proposed synteny with the *Brachypodium* model genome and gene annotation, sequences close to coding regions were used for the development of 70 SNP-based markers. They were mapped on a Arina × Forno Recombinant Inbred Lines population and found to be spread over the complete chromosome 3B. While overall synteny was well maintained, numerous exceptions and inversions of syntenic gene order were identified. Additionally, we found that the majority of recombination events occurred in distal parts of chromosome 3B, particularly in hot-spot regions. Compared with the earlier map based on SSR and RFLP markers, the number of markers increased fourfold. The approach presented here allows fast development of genotype-specific polymorphic markers that can be used for mapping and marker-assisted selection.

Introduction

Bread wheat (*Triticum aestivum* L., $2n = 6x = 42$) is one of the most important food crops in the world. The hexaploid genome of wheat has a size of approximately 17 Gb and consists of the three homoeologous genomes A, B and D. Therefore, most genes are present in three homoeologous copies. Additionally, the wheat genome is very repetitive and contains over 80% of repetitive elements (Bennett and Smith, 1976; Hollister and Gaut, 2009). The characterization of genes involved in agronomically important traits requires the identifications of markers closely linked to the genes. Single-nucleotide polymorphisms (SNPs) are one of the most reliable and reproducible types of sequence-based genetic markers. Sequence information for the hexaploid wheat genome has only recently begun to emerge, and thus, the construction of genetic maps with a high density of genetic markers remains challenging.

Recently, the genomes of rice (*Oryza sativa* L.), sorghum (*Sorghum bicolor* L.) and *Brachypodium distachyon* were sequenced (International Rice Genome Sequencing Project, 2005; Paterson *et al.*, 2009; The International Brachypodium Initiative, 2010). As these genomes have a relatively small size and the gene order in grasses is conserved among different members of the family (Devos, 2005; Gale and Devos, 1998; The International Brachypodium Initiative, 2010), they can be used as models for larger plant genomes. Among the model genomes, *Brachypodium* has the most recent divergence time from wheat —32–39 Million Years Ago (MYA), has the smallest genome size and, therefore, is widely used for gene prediction and construction of synteny-based maps in *Triticeae* (Mayer *et al.*, 2011; Wicker *et al.*, 2009).

During the past few years, wheat genomics has advanced dramatically. The effort to sequence its large genome was divided into smaller steps through construction of physical maps for individual chromosomes. Reducing the complexity of the wheat genome is achieved by isolating single chromosomes or chromosome arms by cytometric flow sorting (Šafář *et al.*, 2010). Nevertheless, there is no reference sequence available for the wheat genome until now. Sequencing of a minimum tiling path of bacterial artificial chromosomes (BAC) clones is under way for several chromosomes (International Wheat Genome Sequencing Consortium www.wheatgenome.org), the most advanced project is the sequencing of chromosome 3B (approximately 1 Gb). A physical map for this chromosome was constructed based on a fingerprinted BAC library with 67 968 clones, genetic mapping and deletion-bin mapping (Paux *et al.*, 2008). Analysis of

collinearity between the physical map of chromosome 3B (cultivar Chinese Spring) and the rice genome has shown that corresponding regions of rice and wheat chromosome 3B are mostly collinear, but there are local rearrangements within this syntenic intervals (Paux *et al.*, 2008). Additionally, based on mapping of expressed sequence tags (EST), a positive gradient of gene density was detected along chromosome 3B from centromere to telomere (Akhunov *et al.*, 2003; Munkvold *et al.*, 2004). According to the study of Akhunov *et al.* (2003), the recombination rate is higher in distal regions of the chromosome. This finding was strongly supported by Choulet *et al.* (2010) based on annotated megabase-sized fragments of sequence and a recent transcriptional map with 3000 loci of chromosome 3B (Rustenholz *et al.*, 2011).

Until now, development of new polymorphic markers was based on very limited resources such as wheat EST data, syntenic information from rice and *Brachypodium* and wheat BAC-sequences from a few loci. New sequencing technologies provide an increase in sequence data from different wheat accessions. SNP assays in combination with new high-throughput genotyping methods are becoming even more attractive for high-resolution genetic mapping and marker-assisted breeding (Agarwal *et al.*, 2008). Several recent studies successfully genotyped a few thousand SNPs using genome survey sequences and EST data (Akhunov *et al.*, 2009, 2010; Allen *et al.*, 2011; Barker and Edwards, 2009; Chao *et al.*, 2009). As modern wheat varieties were bred from a limited number of landraces, the resulting diversity bottleneck lead to modern wheat varieties whose genomes are highly similar (approximately 99.9%) (Chao *et al.*, 2009; Ravel *et al.*, 2006; Trick *et al.*, 2012). For instance, Ravel *et al.* (2006) estimated SNP frequency to be 1 in 335 bp (2.99 SNP/kb), Barker and Edwards (2009) detected 4.29 SNP/kb and Trick *et al.* (2012) found an average density of 1.80 (±1.46) SNP/kb. This low molecular diversity is even more pronounced if only two specific cultivars are compared.

Swiss winter wheat cultivars Arina and Forno were parental lines for a mapping population consisting of recombinant inbred lines (RILs) (Paillard *et al.*, 2003). This population is being used for mapping of different valuable traits, including quantitative trait loci (QTLs) for resistance to several fungal diseases. For QTL mapping, genetic linkage maps for all chromosomes of this population were constructed. The genetic map for chromosome 3B had a length of 199.83 centimorgans (cM) and contained 27 markers.

Here, we report on the production of a SNP database from flow-sorted chromosome 3B of cultivars Arina and Forno. Seventy SNP-based markers were developed and mapped on the Arina x Forno RIL population. Our analysis suggests that while the gene order is overall well conserved between wheat chromosome 3B and chromosome 2 of *Brachypodium*, there are numerous exceptions in microcollinearity. Additionally, we found that the majority of recombination events occur in distal parts of chromosome 3B. The focused approach described in the present study resulted in a fast development of highly specific polymorphic markers that can be applied in high-resolution mapping and marker-assisted selection.

Results

Purification and sequencing of chromosome 3B

Chromosomes 3B of the two Swiss winter wheat cultivars Arina and Forno were purified by flow cytometry to reduce the sample complexity and in particular to avoid the presence of homoeologous sequences from chromosomes 3A and 3D in the DNA samples. The isolated chromosomes 3B of both cultivars were sequenced by Illumina technology. The number of reads obtained for Arina and Forno were 121 931 740 (approximately 9.6 Gb) and 126 662 154 (approximately 10 Gb), respectively. It has been shown that chromosome 3B has a size of 1Gb (Paux *et al.*, 2008). *De novo* assemblies for both cultivars resulted in approximately 300 000 contigs each with an average length of 500 bp. We calculated the average coverage for the contigs containing coding sequences. We produced a graph with the number of contigs plotted against the coverage of those contigs (Figure S1). The peak of the contig number for both cultivars was around 8- to 10-fold coverage (maximum was at eightfold).

Identification of genes based on synteny with *Brachypodium*

To detect putative genes on assembled wheat contigs for both cultivars, we used BLASTN against a database of *B. distachyon* coding sequences. The total number of identified contigs containing coding sequences was 11 563 for Arina and 11 275 for Forno corresponding to approximately 3% of all contigs. This number included different cases. First, if there was exactly one wheat contig matching a specific *Brachypodium* gene, we counted it as a single-copy gene. Second, if two or more different contigs contained homologs of the same *Brachypodium* gene in a different sequence context, they were considered paralogous copies. Finally, if two or more different contigs contained different parts of the same *Brachypodium* gene with only a slight or no overlap, they were considered to represent parts of the same gene. From this data, we defined a 'nonredundant' gene set for chromosome 3B. This gene set represents a list of all *Brachypodium* genes that have homologs in the 3B contigs, regardless if several 3B contigs hit the same *Brachypodium* gene. For example, the coding sequence of *Bradi2g49380* was detected on four different Arina contigs (indicating a gene family with several paralogs), but in the 'nonredundant' set, it was counted only once.

Among all identified gene-containing contigs, we detected 4542 nonredundant genes for Arina and 4293 for Forno. The number of nonredundant genes included all different *Brachypodium* gene hits associated with sequenced wheat contigs. Although both cultivars had a very similar number of nonredundant genes, we detected 249 nonredundant genes more for cultivar Arina. The distal parts of *Brachypodium* chromosome 2 are syntenic to wheat chromosome 3B: *Bd2g00200* – *Bd2g14080* and *Bd2g40150* – *Bd2g62810* (Figure 1a, green bars). About half of the discovered genes had their closest homolog in the syntenic region of *Brachypodium* chromosome 2: 2333 putative genes (51.4%) for Arina and 2131 (49.6%) for Forno. Overall, there are 3674 genes on chromosome 2 of *Brachypodium* in the region syntenic to wheat chromosome 3B. These 3674 genes have potential homologs on chromosome 3B of wheat. Our analysis detected 63.5% and 58% of these genes in the contig sequences of chromosome 3B from cultivar Arina and Forno, respectively. The number of gene sequences that have their best homologs on the nonsyntenic chromosomes of *Brachypodium* was very similar for both cultivars (Table 1). Additionally, as depicted on the heat map (Figure 1a), the localization of these sequences does not show any clustering pattern. This observation suggests that contamination from the flow-sorting procedure was low, and that all wheat chromosomes were equally contributing to the contamination.

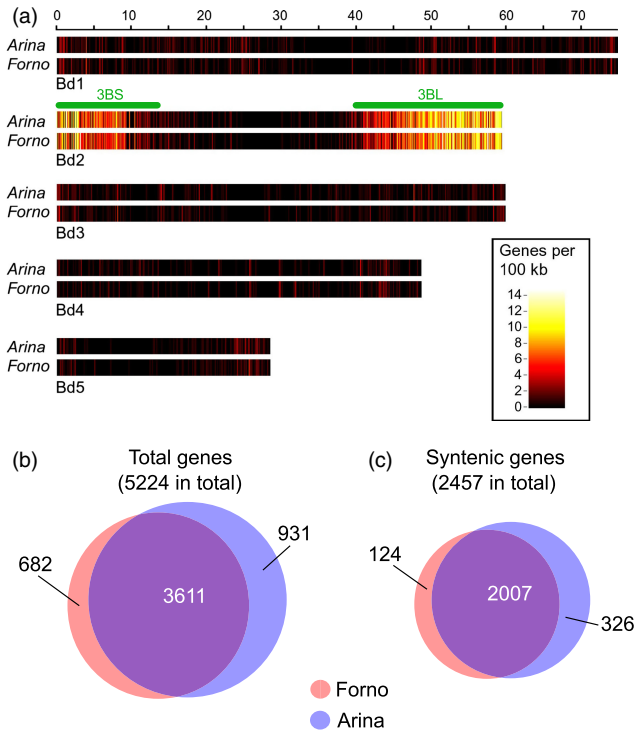

Figure 1 Comparison of identified genes on chromosome 3B of bread wheat cultivars Arina and Forno. (a) Heatmap depicting the distribution of homologs of wheat chromosome 3B genes in the *Brachypodium distachyon* genome. Bd1 through Bd5 correspond to *B. distachyon* chromosomes 1 through 5. The scale at the top is in Mbp. Gene-containing sequence contigs from Arina (top) and Forno (bottom) were mapped independently to the *Brachypodium* genome. The regions syntenic to wheat chromosome 3B have the highest density of homologs. Signals in nonsyntenic regions suggest a presence of true nonsyntenic genes if they occur in both Arina and Forno. Green bars correspond to the syntenic regions of wheat chromosome 3B. (b) Venn diagram with all detected nonredundant coding sequences in each cultivar. (c) Diagram depicts the number of shared and unique genes in the two cultivars with a *Brachypodium* gene homolog from the syntenic region (distal parts of chromosome 2 of *Brachypodium*).

Table 1 Number of wheat chromosome 3B genes on *Brachypodium* chromosomes identified in sequence contigs of cultivars Arina and Forno

Brachypodium chromosome	Arina	Forno
1	764	743
2a	2333	2131
2b	231	218
3	567	560
4	391	391
5	249	244

2a – genes from syntenic part of chromosome 2 of *Brachypodium*.

2b – genes from nonsyntenic region of chromosome 2.

Differences in gene content between Arina and Forno

The combined number of nonredundant genes detected for chromosome 3B of both cultivars was 5224. However, 18% (931) of the genes were present only in Arina and 13% (682 genes) were present only in Forno (Figure 1b). Therefore, approximately 30% of all putative genes detected in the cultivars were different. A very similar result was found for the syntenic genes: 2007 of them were common to both cultivars and 326 putative genes were present only in cultivar Arina, while 124 were present exclusively in cultivar Forno (Figure 1c).

To check whether we can use these data to identify presence/absence polymorphisms in gene content of Arina and Forno, we designed 67 primer pairs from genes specifically found only in Arina or Forno chromosome 3B. We used DNA of wheat cultivars Arina, Forno, Chinese Spring and the Chinese Spring nulli-tetrasomic line (N3BT3D) for amplification of fragments. The last two lines were included as a control for chromosome 3B specificity of the amplicons. Amplification was successful in 45 cases but only five fragments were specific to the 3B chromosome (no amplification from the Chinese Spring nulli-tetrasomic line was obtained, whereas in Chinese Spring, a PCR product was present). The large number of amplified fragments that are not specific for chromosome 3B indicates that these primers also amplified homoeologous sequences. Three of the five chromosome 3B-specific sequences (InDel_g01210, 512-bp-long fragment amplified from Forno, InDel_09780, 143 bp amplified from Arina and InDel_56500 and 818-bp fragment amplified from Forno) were polymorphic between Arina and Forno and were mapped on the genetic map (see below). Therefore, of five fragments specific to chromosome 3B, three were confirmed to be useful as presence/absence polymorphisms.

Identification of SNPs in gene-containing contigs

We chose the Arina *de novo* sequence assembly as a reference on to which Forno Illumina reads were mapped for the identification of polymorphisms. Nucleotide variations between two cultivars were identified using the CLC Assembly Cell (Figure 2). In total, 1 835 214 SNPs were detected. To reduce the problem of unspecific comparisons mostly in repetitive DNA regions, only contigs containing coding sequences of genes were chosen for further analysis. Gene-containing contigs were selected for further analysis if they had a unique gene hit or if a contig did not have any overlaps with other contigs, which had the same gene hit. This selection step was made to exclude paralogous gene copies for marker development.

A further selection among the contigs containing suitable coding sequences and SNPs was made to reduce the number of SNPs caused by sequencing errors. First, we applied a sequence coverage threshold for each candidate SNP. Nucleotide polymorphisms in regions with a coverage below five reads were not considered for SNP marker development (Figure 2). Second, we mapped all Arina reads to the reference Arina assembly to control base calling and to eliminate candidate SNP positions with low sequence coverage. All SNPs with a coverage below five in Arina or ambiguous base calling were excluded. In total, 737 contigs with high-quality SNPs were selected. Among them, 448 contigs had gene hits from nonsyntenic regions of *Brachypodium* and 289 were from the region which corresponded to chromosome 3B. From them, 131 belonged to coding sequences from the short chromosome arm and 158 contigs had gene hits corresponding to the long arm of 3B.

Development of a cultivar-specific genetic map of chromosome 3B based on 70 SNP markers

For SNP mapping, we used only contigs representing wheat genes, which have their homologs in the 3B syntenic regions of

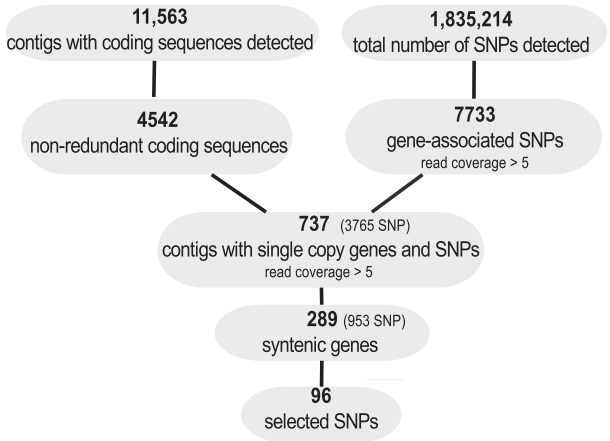

Figure 2 Scheme of SNP selection steps. Coding sequences (CDS) on the contigs of Arina assembly were detected using a database of *Brachypodium distachyon* coding sequences. The number of nonredundant coding sequences included all different *Brachypodium* gene hits associated with sequenced wheat contigs.

Brachypodium. We chose genes expected to be regularly spread across the whole chromosome based on syntenic location in *Brachypodium*. In total, 96 SNPs were selected for the genetic mapping on 178 RILs derived from a cross between cultivar Arina and cultivar Forno (Paillard *et al.*, 2003).

A KASPar assay was chosen for genotyping of the 178 RIL plants. Before designing primers for the KASPar assay, we masked the coding parts of the selected contigs. Only SNPs located upstream or downstream of the gene's CDS were selected. This strategy aimed at reducing the possibility that homoeologous or paralogous copies of targeted genes were amplified. As a result, 87 (out of 96) SNP containing fragments were amplified, among them 64 were polymorphic and suitable for genetic mapping. In addition to these 64 markers, six polymorphic SNPs from eight tested were mapped on the RIL population in a pilot experiment to test the feasibility of the assay, resulting in a total of 70 markers derived from the SNP assay. Therefore, in total, 104 SNP markers were tested, 95 of them amplified and 70 markers were mapped (Table S1).

The initial genetic map of chromosome 3B of Arina x Forno RIL population consisted of 27 markers and had a size of 199.83 cM. After the integration of 70 additional SNP and three InDel markers, the length of the map was 202.2 cM (Figure 3). This indicated that the integration of the added markers was of good quality. The names of new SNP-based markers were chosen according to the *Brachypodium* gene homologs, which were present on the wheat contigs that contained the SNPs (e.g. *Brachypodium* gene *Bd2g47720* – wheat marker g47720).

To test whether the SNP markers specifically developed from Arina and Forno were applicable in a larger set of genetic material, we tested a subset of 48 SNP markers on an additional 44 wheat varieties with different origins: European spring and winter wheat germplasm, US germplasm, Mexican and Canadian varieties and Chinese landraces (Table S2). The test revealed a presence of both Arina and Forno alleles with variable frequencies from 0.02 to 1 for each allele (Figure S2). Among the 48 SNPs, only one SNP represented a private allele of cv. Forno, whereas all other SNPs were present in a least one of the tested lines. Thus, the developed SNPs are widely polymorphic in germplasm and add to the available SNP markers in wheat.

SNP mapping reveals overall good collinearity with *Brachypodium* with several exceptions

We expected to find a good conservation between gene order on chromosome 2 of *Brachypodium* and wheat chromosome 3B. Indeed, the marker order on the newly constructed wheat map was mostly consistent with the gene order found on chromosome 2 of *Brachypodium*. Overall, only eight of the 73 (70 SNP and three InDel markers) analysed markers were in nonsyntenic positions. As depicted in Figure 3, we observed co-segregation of a few groups of markers, especially in the centromeric region. With a few exceptions, these groups were formed by markers with their *Brachypodium* gene homologs in syntenic positions. For example, the largest group of 12 co-segregating markers included only one nonsyntenic marker, g62500 (Figure 3, cluster B in cyan). The 11 syntenic genes in this cluster (*Bd2g07680* – *Bd2g09470*) span a region of 1.73 Mb in *Brachypodium*. In wheat, the corresponding gene-based SNP markers co-segregated on the genetic map in a region close to the centromere.

We detected four inversions on the genetic map of chromosome 3B by comparison with the syntenic order of the genes on *Brachypodium* chromosome 2 (Figure 3, clusters A, C, D, E). The sizes of the identified inversions ranged from 0.6 cM and 1.3 cM in proximal regions to 7.6 cM in the telomeric part. One of these inversions (cluster C) consisted of nine genes (*Bd2g46680* – *Bd2g47750*) located in a 1-Mb interval on *Brachypodium* chromosome 2 (Figure 3, cluster C in cyan). In wheat, this inversion comprised a genetic interval of 1.2 cM. Such information on breaks in synteny and local, small-scale inversions will be essential for high-resolution mapping projects.

SNPs are unevenly distributed along chromosome 3B

To analyse the distribution of gene-associated SNP markers on the genetic map of chromosome 3B, we examined the correlation between the distribution of the 289 high-quality, syntenic, gene-associated SNPs and the distribution of all syntenic *Brachypodium* gene homologs detected on wheat contigs (2333 Arina genes) on chromosome 2 of *Brachypodium*. We mapped *in silico* all identified syntenic genes from wheat contigs to their homologous positions on the chromosome 2 of *Brachypodium*. The correlation was tested by calculating the ratio between the number of selected genes with SNPs and the total number of identified syntenic genes. The ratio was computed using a sliding window approach: we took intervals of 1.2 Mb with incremental steps of 100 kb (Table S3). The ratios, obtained for each interval, are shown in Figure 3 (blue graph). This analysis revealed that the type of SNPs studied here were distributed unevenly: the number of genes with SNPs, normalized to the total amount of genes, had a few peaks and was generally higher in the region corresponding to the short arm of chromosome 3B. This suggests that the frequency of gene-associated SNPs in cultivars Forno and Arina is higher in 3BS than in 3BL.

Low recombination in the centromeric region of chromosome 3B

We also wanted to compare the genetic distances on the map of wheat chromosome 3B with the physical size and number of syntenic genes in the corresponding chromosomal regions in *Brachypodium.* We focused our analysis on a region of 18 cM that included the centromeric region of the map with a large gene cluster (Figure 3, purple block). There were a total of 47 mapped gene-associated SNPs and 1255 syntenic *Brachypodium*

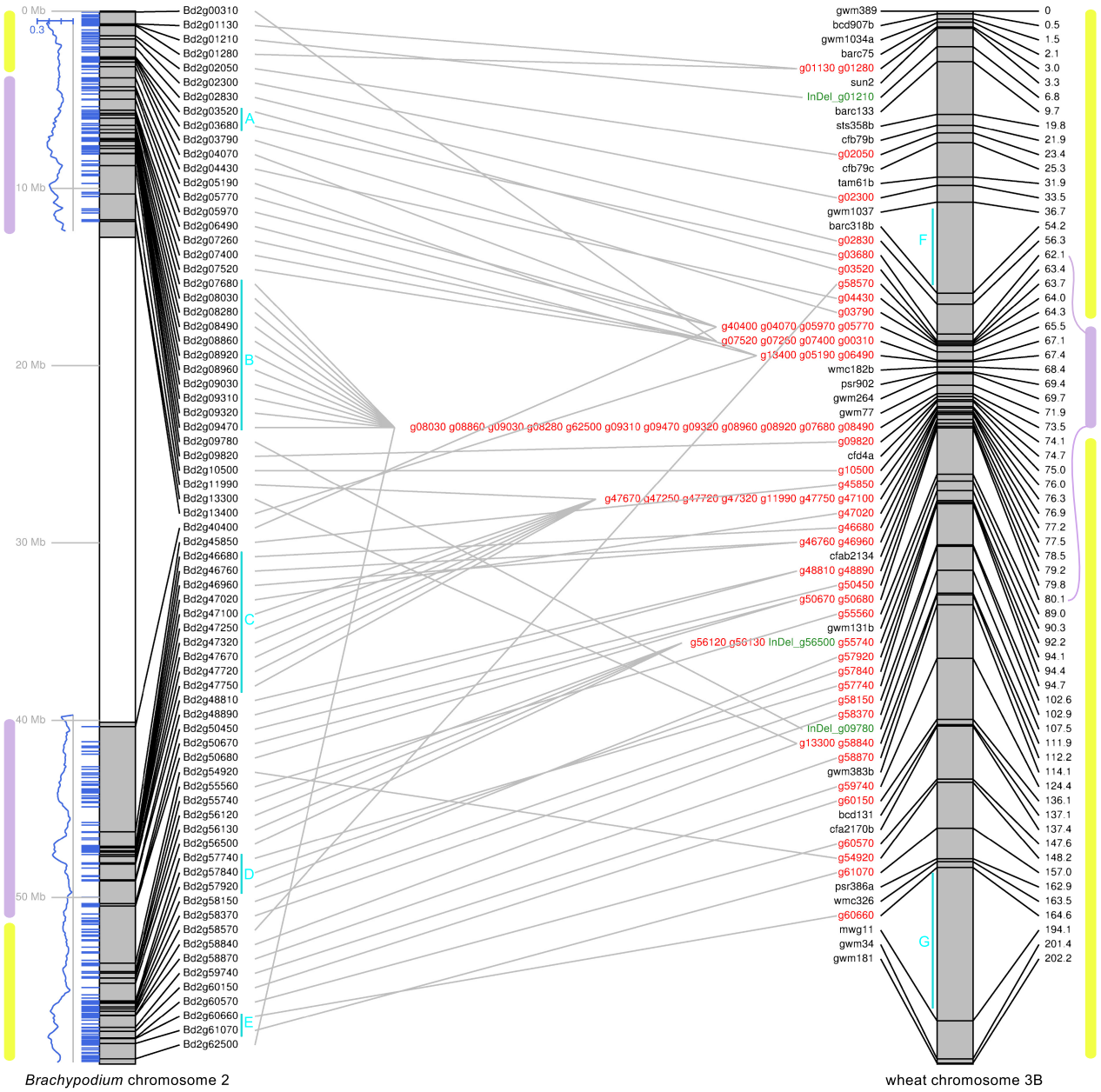

Figure 3 Comparison of the genetic map of wheat chromosome 3B of RIL Arina x Forno population and the physical map of *Brachypodium* chromosome 2. In grey: syntenic regions of *Brachypodium* chromosome 2 and wheat chromosome 3B. In red: newly mapped gene-based SNP markers. Marker names correspond to the gene homologs in *Brachypodium*. Genetic distances are given in centimorgans (cM). Blue bars: physical positions of *Brachypodium* gene homologs corresponding to the selected gene-associated contigs with SNPs of wheat chromosome 3B; gene homologs that were mapped as SNP markers are represented both as a blue bar and black bar on the *Brachypodium* chromosome. Blue graph depicts a ratio between the number of selected wheat gene-containing contigs with SNPs and the total number of identified nonredundant *Brachypodium* genes in wheat sequenced contigs. The ratio was calculated using a sliding window of 1.2 Mb and incremental step 100 kb. In cyan: A, C, D, E – regions of detected inversions between wheat and *Brachypodium,* B is a cluster of mapped *Brachypodium* genes, regions F and G – gaps on genetic map. In purple and yellow: centromeric region of genetic map (purple) and corresponding physical region on *Brachypodium* chromosome 2; distal parts of genetic map with corresponding physical *Brachypodium* regions (yellow). In green: marker derived from gene presence–absence polymorphism.

genes detected in this region based on the sequences from wheat cultivar Arina. The wheat region of 18 cM around the centromere corresponded to a segment on *Brachypodium* chromosome 2, which spanned a total size of 19.34 Mb, with 9.23 Mb on the syntenic region of the short arm and 10.11 Mb on the syntenic region of the long arm. Therefore, in this centromeric region, 53.79% of all syntenic genes (1255 of 2333) were found in a region corresponding to 8.9% (18 cM) of the total genetic distance in wheat and 62.89% of the total physical size 30.75 Mb of the syntenic region in *Brachypodium*.

Brachypodium genes homologous to the gene-associated SNP markers in the distal parts of the genetic map were assigned to

the corresponding yellow blocks (Figure 3). Physical intervals corresponding to the distal parts of the genetic map had sizes of 2.63 Mb for the short arm of chromosome 3B and 8.78 Mb for the long arm of chromosome 3B (together corresponding to 37.11% of the 30.75 Mb physical region syntenic in *Brachypodium*). The two distal parts of the genetic map comprised 91.1% of the total map size in wheat (184.2 cM) and contained 46.21% (1078) of detected genes. These observations demonstrate that in Arina x Forno population, there is a strong suppression of recombination in the centromeric region of wheat chromosome 3B, but also that there are many genes in this region.

Gaps on the genetic map

The Arina/Forno SNP map contained two relatively large regions not covered by genetic markers (Figure 3, region F and region G). The region G (g60660 – mwg11) has a size of 29.5 cM and the last gene-associated SNP mapped to this area (g60660) is 36.7 cM away from the telomeric end of the genetic map. The number of detected syntenic *Brachypodium* gene homologs in the Arina dataset was 115 for the region from *Bd2g61070* – end of the *Brachypodium* chromosome 2. Therefore, further mapping could probably place some of these 115 genes in this gap. Additionally, Saintenac *et al.* (2009) showed that recombination frequency in this region is very high. Furthermore, other genetic maps indicated a similar absence of markers: on the ITMI composite map, marker wmc326 (on our map, it is placed 1.1 cM proximal from the described interval) is located at position 114 cM, while the total size of the map is 148 cM (Somers *et al.*, 2004). On the map of double haploid lines Avalon x Cadenza, wmc326 was mapped to position 138.8 cM (total size of the map was 183.0 cM). In our case, wmc326 is located at 163.5 cM position (total size of the map is 202.2 cM). Gaps are representing the regions where genetic distance is large (a high level of recombination), but the number of markers is small. Based on the syntenic comparison, there is no evidence for large physical regions in the corresponding regions of wheat. Therefore, the described gap g60660–mwg11 can most likely be explained by a high recombination level in this region.

The region F (gwm1037 – barc318b) is 17.5 cM long and does not contain any genetic markers. We checked the number of *Brachypodium* gene homologs, which were detected in Arina in the interval *Bd2g02300 – Bd2g02830 (*g02300 is 3.2 cM distal from gwm1037 and g02830 is 2.1 cM proximal from barc318b). There are 31 *Brachypodium* genes detected. Hence, this gap can possibly also be explained by the presence of a recombination hot-spot or by nonsyntenic genes intercalated in this region.

Discussion

Gene-associated SNP markers based on whole chromosome sequencing are highly specific

Mapping in hexaploid wheat is complicated by the difficulty to derive specific markers because of the highly repetitive genome and the presence of both paralogous and homoeologous copies of a particular gene in the three subgenomes. Therefore, an essential part of any mapping strategy is to increase marker specificity. In the present study, we used sequences from isolated chromosome 3B. Flow cytometric analysis of DAPI-stained mitotic chromosomes of wheat cv. Chinese Spring results in a histogram of relative fluorescence intensity (flow karyotype) comprising three composite peaks of groups of chromosomes and a peak corresponding to chromosome 3B (Vrána *et al.*, 2000). Therefore, flow sorting of chromosome 3B is easier than the isolation of other wheat chromosomes. These can be sorted as chromosome arms from cytogenetic stocks in which the arms are stably maintained as telocentric chromosomes (Doležel *et al.*, 2007). Alternatively, fractions highly enriched for target chromosomes and free of homoeologous chromosomes can be obtained by applying narrow sort windows (J. Doležel, unpublished data). Any of these strategies reduces DNA sample complexity, which facilitates efficient and targeted development of genetic markers (Wenzl *et al.*, 2010).

The main strategy of our study was to use gene-associated SNPs to improve marker specificity for chromosome 3B and to get an even distribution of markers based on gene synteny with *Brachypodium*. To gain specificity for designed SNP markers both between homoeologous chromosomes and within chromosome 3B, we used a set of criteria for the selection of SNP. All repetitive sequences were excluded from the analysis, because amplification from repetitive regions could lead to unspecific amplification from similar repetitive sequences. Furthermore, a limitation of the sequence read coverage threshold for each SNP position was established. We selected only contigs with nonredundant gene hits and analysed SNPs located outside of the coding region. This setup resulted in a low number of nonspecific markers and a final success rate of approximately 74% (70 of 95) for SNP marker mapping. We used very stringent parameters for SNP selection (e.g. only syntenic genes, only one contig per given gene, read coverage >5, the SNP had to be present in all reads). It is therefore possible that more flexible parameters might give a higher yield of SNPs for mapping. Nevertheless, we consider the number of identified SNPs and the yield in marker development a success, because low polymorphism is a general problem in wheat genetics.

Several prior studies showed that wheat has a low level of varietal SNP polymorphisms useful for marker development (Barker and Edwards, 2009). A recent study by Allen *et al.* (2011) revealed that among 1659 SNPs selected using a cDNA library of five wheat varieties, 67% SNPs were polymorphic between different varieties, but only 10% of polymorphic SNPs showed genome specificity. The approach for optimization of genome specificity, suggested in the study by Allen *et al.* (2011), included re-design of KASPar probes based on alignments of three genome-specific sequences for corresponding regions from the Chinese Spring fivefold sequence database (www.cerealsdb.uk.net). They report that their approach indeed increased genome specificity of SNPs, but the method is time consuming and laborious. Trick *et al.* (2012) mapped 39 SNPs on tetraploid wheat using RNAseq and wheat unigenes. Discrimination between the copies on the two different genomes was done by alignment of sequences to unigenes and the design of genome-specific primers for KASPar, which led to a success rate of 74% in this study.

The strategy used in our study resulted in high percentage of polymorphic SNP markers. The vast majority of produced SNP markers (66 of 70) were genome specific. This level of efficiency is high given the problem of three closely related homoeologous genomes in wheat and was obtained without any additional steps such as genomic alignments and re-designing of primers for the KASPar assay. Furthermore, preselection of potential SNPs based on syntenic gene homologs allowed us to obtain a more predictable distribution of SNP markers on the genetic map. Preselection could be particularly advantageous for high-resolution mapping projects.

Wheat and *Brachypodium* show a high degree of synteny but also many exceptions

We used a synteny-based approach to identify putative genes on sequenced wheat contigs. Because *B. distachyon* is the closest wheat relative with a small and entirely sequenced genome, identification of wheat genes and their order based on the synteny with *Brachypodium* have been used extensively (Bossolini *et al.*, 2007; Quraishi *et al.*, 2011). Our analysis revealed that about half of the identified genes (2333 from 4542 in Arina and 2131 from 4293 in Forno) had their homologs in distal parts of chromosome 2 of *Brachypodium*, the 3B syntenic regions. Additionally, the absence of large gene clusters with corresponding gene homologs on other *Brachypodium* chromosomes indicated an absence of large chromosome translocations. However, there is still a possibility of large translocations in gene-poor regions that cannot be excluded with our approach. Genetic mapping of gene-associated SNPs also revealed several inversions between gene order on *Brachypodium* chromosome 2 and the genetic map of chromosome 3B (Figure 3, cyan bars).

Gene content of chromosome 3B

Besides the syntenic genes, a large number of nonsyntenic genes were identified. This observation is consistent with previous findings about gene content of wheat chromosome 3B. It was suggested that nonsyntenic genes in wheat are interspersed within a very conserved syntenic gene set (Choulet *et al.*, 2010). Indeed, we found that roughly 50% of the detected genes mapped to nonsyntenic regions in *Brachypodium* in both Arina and Forno assemblies. Several recent studies provided estimates of the total gene number for chromosome 3B between 6000 and 8400 genes (Choulet *et al.*, 2010; Paux *et al.*, 2006). Our observations are slightly different: we detected roughly 11 000 gene-containing contigs in each cultivar. However, because we did not conduct a detailed gene annotation, this number possibly includes many pseudogenes, gene fragments and misassembled parts of the same genes. Nevertheless, we were able to identify a minimal gene set for chromosome 3B consisting of approximately 3600 nonredundant genes (Figure 1b). This minimal set of genes was detected independently in both Arina and Forno datasets, suggesting a high accuracy of gene prediction for this set.

In Arina, 14% of genes with synteny to *Brachypodium* chromosome 2 were cultivar specific, in Forno, 5.8%. These numbers could be explained by the presence of real gene deletions in Arina and Forno genomes or by insufficient sequence coverage of the samples. We could confirm some presence/absence polymorphisms by mapping of InDel markers, but most of the other amplification products were not chromosome 3B specific, probably due to homoeologous sequences in the A and D genomes. The calculated coverage for our data set is eightfold, and we expect 99.97% of the genes to be sampled (Lander and Waterman, 1988). However, a purely random coverage cannot be expected: a certain bias in the sequence coverage has been shown to be caused by multiple displacement amplification (MDA) of flow-sorted chromosomes and the sequencing itself (Pinard *et al.*, 2006; Wicker *et al.*, 2011). Additionally, the high number of cultivar-specific genes could be caused by our gene annotation procedure, which was based on best homology to Brachypodium genes. For example, it is possible that sequence contigs from Arina and Forno cover different parts of a given gene. Homology search in Brachypodium could then result in two different genes as best hit, wrongly indicating that both genes are cultivar specific.

The level of recombination varies strongly along the chromosome

The majority of mapped markers were placed in the proximal centromeric region of chromosome 3B. In addition, numerous markers from the centromeric region co-segregated genetically. The proximal centromeric region of the genetic map spanned 18 cM (<10% of the genetic map) and included 1255 genes in the syntenic regions of Brachypodium (53.8% of the genes from the syntenic regions of *Brachypodium* chromosome 2). The rest of chromosome 3B (distal regions) had a genetic size of 62.1 cM and 122.1 cM and included 210 (9.0%) and 868 (37.2%) of the genes in the syntenic regions of chromosome 2, respectively. This observation demonstrates that recombination was highly suppressed in the centromeric region. This finding is consistent with previous studies: Saintenac *et al.* (2009) analysed crossover frequency using data from physical and genetic maps of wheat chromosome 3B. They detected a crossover frequency of 90% in distal subtelomeric parts and a very low recombination in the proximal centromeric region of chromosome 3B. The comparison of the chromosome 3B genetic map and gene content as estimated by comparison with syntenic chromosome 2 of *Brachypodium* very well reflected the detected low recombination rate around the centromeric region (Figure 3, purple and yellow blocks). This fact suggests that mapping in this region is especially challenging.

Differences in gene content between two cultivars: contamination or tool for marker development?

The absence of large clusters of *Brachypodium* gene homologs elsewhere than in the syntenic regions of *Brachypodium* chromosome 2 indicated that the contamination of the chromosome 3B fraction with other chromosomes during flow cytometric sorting was low. Furthermore, a maximal degree of contamination could be estimated based on a total number of genes exclusively present in Arina or in Forno (Figure 1b). The detected number of cultivar-specific genes in Arina and Forno was 931 and 682, respectively. If we assume that identification of those genes was caused entirely by contamination during the flow sorting, the maximal estimation would be 20% of contaminated fraction for Arina and 16% for Forno. This is more than estimated by fluorescent *in situ* hybridization (FISH) on fractions of sorted chromosomes (3.2% and 8.9% for Arina and Forno, respectively). However, if the detected variation in gene content between cultivars Arina and Forno was based not solely on contamination, it has at least two potential applications. First, genes that are absent in one cultivar and present in another could be used as dominant presence/absence markers for genetic mapping. This concept is supported by the fact that we could map 3 of 5 chromosome 3B-specific gene-based InDel markers (Figure 3, InDels in green). However, the development of such markers is inefficient as the primers developed from presence/absence polymorphism are mostly not chromosome specific and also amplify homoeologous sequences. A second application of InDel variation could be in candidate gene selection in fine mapping studies.

Whole chromosome sequence-based mapping in wheat

The strategy based on sequencing flow-sorted chromosomes can be used for targeted genotyping of different wheat varieties or populations both in high and low throughput manner. We

selected SNPs that were gene associated, but not positioned in the coding sequences themselves. This factor potentially provides a higher rate of polymorphisms between elite wheat varieties in comparison with SNPs in coding regions. The developed SNP markers are genome specific and therefore can be used not only for screening of homozygous plants but also applicable for mapping in heterozygous populations. The analysis of a subset of 48 SNP-based markers, specifically developed for cultivars Forno and Arina, in a broader set of 44 genotypes allowed us to determine the frequency of polymorphisms in such germplasm. The results suggest that developed SNP can be widely used in the genepool. In conclusion, SNP markers are becoming increasingly important for wheat genetic mapping and marker-assisted selection. The strategy to develop SNP markers applied in the present study is a fast and efficient approach to genetic mapping. A combination of sequencing of individual chromosomes or chromosome groups and selection of SNPs associated with genes but located outside of the coding sequences produced highly specific results.

Experimental procedures

Plant material

Two varieties of winter wheat (*Triticum aestivum* L., 2n=6x=42) 'Arina' and 'Forno' were chosen for SNP analysis. Genetic mapping was carried out on a population of 178 RILs, derived from a cross of cultivar 'Arina' and cultivar 'Forno' (Paillard *et al.*, 2003). DNA of cultivars 'Arina', 'Forno', 'Chinese Spring' and nulli-tetrasomic chromosome 3B line N3BT3D of 'Chinese Spring' was used for InDel PCR analysis. DNA extraction from leaf material was carried out as described in Stein *et al.* (2001).

Flow sorting of chromosome 3B

Purification of chromosome 3B was carried out by flow sorting. Liquid suspensions of mitotic chromosomes were prepared from synchronized root tips of both wheat cultivars according to Vrána *et al.* (2000). Chromosome samples were stained by DAPI, and chromosome 3B was isolated by flow cytometric sorting as described by Kubaláková *et al.* (2002). The chromosomes were sorted in several batches of 25 000 chromosomes with the purity of 96.8% and 91.1% for Arina and Forno, respectively, as determined by FISH with probes for GAA and *Afa* repeat. DNA of isolated chromosomes was amplified *via* MDA using Illustra GenomiPhi V2 DNA amplification kit (GE Healthcare). Two and four independent amplifications were made for Arina and Forno, respectively. In total, 8.61 μg and 10.92 μg DNA were obtained for Arina and Forno, respectively.

Sequencing and assembly of chromosomes 3B of cultivars Arina and Forno

Sequencing of chromosomes was performed by GATC Biotech, Konstanz, Germany. For each cultivar, 5 μg DNA was sequenced by Illumina technology with independent libraries of 200- to 250-bp sequence fragments and paired ends. After quality trimming, 100-bp-long Illumina reads of Arina and Forno were separately assembled *de novo* using 'clc_novo_assemble' command with default parameters of CLC Assembly Cell 3.22 software (CLC bio, Aarhus, Denmark).

Calculation of expected gene coverage

To calculate the expected gene coverage, we used the following formula based on the Lander–Waterman approach:

$P = 1-e^{-c}$, where P is a probability that any base in contigs with genes is covered and c – average coverage of the contigs with coding sequences. The average coverage was calculated as described above (Figure S1).

Identification and selection of SNPs

Forno reads were mapped to the Arina *de novo* assembly using function 'clc_ref_assemble_long' of CLC Assembly Cell. Potential SNPs were identified using function 'find_variations' of CLC Assembly Cell. Read coverage for each base pair position in targeted contigs was extracted using 'assembly_info' CLC Assembly Cell. BLASTN against *B. distachyon* coding sequences database was used to detect putative genes on wheat contigs. Perl scripts were written to mine and extract the data for SNP selection and are available upon request.

RIL genotyping and construction of the genetic map

One hundred and seventy-eight RILs Arina x Forno were chosen for testing of 104 identified SNPs. Genotyping was performed by KBioscience (Herts, UK) using fluorescence-based competitive allele-specific PCR (KASPar) assay. The genetic map was constructed using MultiPoint software (Mester *et al.*, 2003). Distances were calculated using Kosambi function (Kosambi, 1944).

Acknowledgements

We thank Bea Senger for technical assistance with the plant material, Gerhard Herren for the support with the PCR analysis and Roi Ben David for the advice on construction of the genetic map. The research leading to these results received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under the grant agreement FP7-212019 (Triticeae Genome) and was supported in the framework of the European Cooperation in Science and Technology FA0604 (Tritigen). This work was also supported by a grant from the Swiss National Science Foundation 31003A_127061.

References

Agarwal, M., Shrivastava, N. and Padh, H. (2008) Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep.* **27**, 617–631.

Akhunov, E.D., Goodyear, A.W., Geng, S., Qi, L.-L., Echalier, B., Gill, B.S., Miftahudin, Gustafson, J.P., Lazo, G., Chao, S., Anderson, O.D., Linkiewicz, A.M., Dubcovsky, J., Rota, M.L., Sorrells, M.E., Zhang, D., Nguyen, H.T., Kalavacharla, V., Hossain, K., Kianian, S.F., Peng, J., Lapitan, N.L.V., Gonzalez-Hernandez, J.L., Anderson, J.A., Choi, D.-W., Close, T.J., Dilbirligi, M., Gill, K.S., Walker-Simmons, M.K., Steber, C., McGuire, P.E., Qualset, C.O. and Dvorak, J. (2003) The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. *Genome Res.* **13**, 753–763.

Akhunov, E., Nicolet, C. and Dvorak, J. (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theor. Appl. Genet.* **119**, 507–517.

Akhunov, E.D., Akhunova, A.R., Anderson, O.D., Anderson, J.A., Blake, N., Clegg, M.T., Coleman-Derr, D., Conley, E.J., Crossman, C.C., Deal, K.R., Dubcovsky, J., Gill, B.S., Gu, Y.Q., Hadam, J., Heo, H., Huo, N., Lazo, G.R., Luo, M.-C., Ma, Y.Q., Matthews, D.E., McGuire, P.E., Morrell, P.L., Qualset, C.O., Renfro, J., Tabanao, D., Talbert, L.E., Tian, C., Toleno, D.M., Warburton, M.L., You, F.M., Zhang, W. and Dvorak, J. (2010) Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. *BMC Genomics*, **11**, 702.

Allen, A.M., Barker, G.L.A., Berry, S.T., Coghill, J.A., Gwilliam, R., Kirby, S., Robinson, P., Brenchley, R.C., D'Amore, R., McKenzie, N., Waite, D., Hall, A.,

Bevan, M., Hall, N. and Edwards, K.J. (2011) Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat *(Triticum aestivum* L.*). Plant Biotechnol. J.* **9**, 1086–1099.

Barker, G.L.A. and Edwards, K.J. (2009) A genome-wide analysis of single nucleotide polymorphism diversity in the world's major cereal crops. *Plant Biotechnol. J.* **7**, 318–325.

Bennett, M.D. and Smith, J.B. (1976) Nuclear DNA amounts in angiosperms. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **274**, 227–274.

Bossolini, E., Wicker, T., Knobel, P.A. and Keller, B. (2007) Comparison of orthologous loci from small grass genomes *Brachypodium* and rice: implications for wheat genomics and grass genome annotation. *Plant J.* **49**, 704–717.

Chao, S., Zhang, W., Akhunov, E., Sherman, J., Ma, Y., Luo, M.-C. and Dubcovsky, J. (2009) Analysis of gene-derived SNP marker polymorphism in US wheat *(Triticum aestivum* L.*)* cultivars. *Mol. Breed.* **23**, 23–33.

Choulet, F., Wicker, T., Rustenholz, C., Paux, E., Salse, J., Leroy, P., Schlub, S., Paslier, M.-C.L., Magdelenat, G., Gonthier, C., Couloux, A., Budak, H., Breen, J., Pumphrey, M., Liu, S., Kong, X., Jia, J., Gut, M., Brunel, D., Anderson, J.A., Gill, B.S., Appels, R., Keller, B. and Feuillet, C. (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. *Plant Cell*, **22**, 1686–1701.

Devos, K.M. (2005) Updating the 'crop circle'. *Curr. Opin. Plant Biol.* **8**, 155–162.

Doležel, J., Kubaláková, M., Paux, E., Bartoš, J. and Feullet, C. (2007) Chromosome-based genomics in the cereals. *Chromosome Res.* **15**, 51–66.

Gale, M.D. and Devos, K.M. (1998) Comparative genetics in the grasses. *Proc. Natl Acad. Sci. USA*, **95**, 1971–1974.

Hollister, J.D. and Gaut, B.S. (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* **19**, 1419–1428.

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.

Kosambi, D. (1944) The estimation of map distances from recombination values. *Ann. Eugen.* **12**, 172–175.

Kubaláková, M., Vrána, J., Číhalíková, J., Šimková, H. and Doležel, J. (2002) Flow karyotyping and chromosome sorting in bread wheat *(Triticum aestivum* L.*). Theor. Appl. Genet.* **104**, 1362–1372.

Lander, E.C. and Waterman, M.S. (1988) Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics*, **2**, 231–239.

Mayer, K.F.X., Martis, M., Hedley, P.E., Šimková, H., Liu, H., Morris, J.A., Steuernagel, B., Taudien, S., Roessner, S., Gundlach, H., Kubaláková, M., Suchánková, P., Murat, F., Felder, M., Nussbaumer, T., Graner, A., Salse, J., Endo, T.R., Sakai, H., Tanaka, T., Itoh, T., Sato, K., Platzer, M., Matsumoto, T., Scholz, U., Doležel, J., Waugh, R. and Stein, N. (2011) Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell*, **23**, 1249–1263.

Mester, D., Ronin, Y., Minkov, D., Nevo, E. and Korol, A. (2003) Constructing large-scale genetic maps using an evolutionary strategy algorithm. *Genetics*, **165**, 2269–2282.

Munkvold, J.D., Greene, R.A., Bermudez-Kandianis, C.E., Rota, C.M.L., Edwards, H., Sorrells, S.F., Dake, T., Benscher, D., Kantety, R., Linkiewicz, A.M., Dubcovsky, J., Akhunov, E.D., Dvorák, J., Miftahudin, Gustafson, J.P., Pathan, M.S., Nguyen, H.T., Matthews, D.E., Chao, S., Lazo, G.R., Hummel, D.D., Anderson, O.D., Anderson, J.A., Gonzalez-Hernandez, J.L., Peng, J.H., Lapitan, N., Qi, L.L., Echalier, B., Gill, B.S., Hossain, K.G., Kalavacharla, V., Kianian, S.F., Sandhu, D., Erayman, M., Gill, K.S., McGuire, P.E., Qualset, C. O. and Sorrells, M.E. (2004) Group 3 chromosome bin maps of wheat and their relationship to rice chromosome 1. *Genetics*, **168**, 639–650.

Paillard, S., Schnurbusch, T., Winzeler, M., Messmer, M., Sourdille, P., Abderhalden, O., Keller, B. and Schachermayr, G. (2003) An integrative genetic linkage map of winter wheat *(Triticum aestivum* L.*). Theor. Appl. Genet.* **107**, 1235–1242.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A.K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I.V., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Otillar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Rahman, M., Ware, D., Westhoff, P., Mayer, K.F.X., Messing, J. and Rokhsar, D.S. (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, **457**, 551–556.

Paux, E., Roger, D., Badaeva, E., Gay, G., Bernard, M., Sourdille, P. and Feuillet, C. (2006) Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. *Plant J.* **48**, 463–474.

Paux, E., Sourdille, P., Salse, J., Saintenac, C., Choulet, F., Leroy, P., Korol, A., Michalak, M., Kianian, S., Spielmeyer, W., Lagudah, E., Somers, D., Kilian, A., Alaux, M., Vautrin, S., Bergès, H., Eversole, K., Appels, R., Šafář, J., Šimková, H., Doležel, J., Bernard, M. and Feuillet, C. (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. *Science*, **322**, 101–104.

Pinard, R., de Winter, A., Sarkis, G.J., Gerstein, M.B., Tartaro, K.R., Plant, R.N., Egholm, M., Rothberg, J.M. and Leamon, J.H. (2006) Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics*, **7**, 216–237.

Quraishi, U.M., Murat, F., Abrouk, M., Pont, C., Confolent, C., Oury, F.X., Ward, J., Boros, D., Gebruers, K., Delcour, J.A., Courtin, C.M., Bedo, Z., Saulnier, L., Guillon, F., Balzergue, S., Shewry, P.R., Feuillet, C., Charmet, G. and Salse, J. (2011) Combined meta-genomics analyses unravel candidate genes for the grain dietary fiber content in bread wheat *(Triticum aestivum* L.*). Funct. Integr. Genomics*, **11**, 71–83.

Ravel, C., Praud, S., Murigneux, A., Canaguier, A., Sapet, F., Samson, D., Balfourier, F., Dufour, P., Chalhoub, B., Brunel, D., Beckert, M. and Charmet, G. (2006) Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat *(Triticum aestivum* L.*). Genome*, **49**, 1131–1139.

Rustenholz, C., Choulet, F., Laugier, C., Šafář, J., Šimková, H., Doležel, J., Magni, F., Scalabrin, S., Cattonaro, F., Vautrin, S., Bellec, A., Bergès, H., Feuillet, C. and Paux, E. (2011) A 3,000-loci transcription map of chromosome 3B unravels the structural and functional features of gene islands in hexaploid wheat. *Plant Physiol.* **157**, 1596–1608.

Šafář, J., Šimková, H., Kubaláková, M., Číhalíková, J., Suchánková, P., Bartoš, J. and Doležel, J. (2010) Development of chromosome-specific BAC resources for genomics of bread wheat. *Cytogenet. Genome Res.* **129**, 211–223.

Saintenac, C., Falque, M., Martin, O.C., Paux, E., Feuillet, C. and Sourdille, P. (2009) Detailed recombination studies along chromosome 3B provide new insights on crossover distribution in wheat *(Triticum aestivum* L.*). Genetics*, **181**, 393–403.

Somers, D.J., Isaac, P. and Edwards, K. (2004) A high-density microsatellite consensus map for bread wheat *(Triticum aestivum* L.*). Theor. Appl. Genet.* **109**, 1105–1114.

Stein, N., Herren, G. and Keller, B. (2001) A new DNA extraction method for high-throughput marker analysis in a large-genome species such as *Triticum aestivum*. *Plant Breed.* **120**, 354–356.

The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, **463**, 763–768.

Trick, M., Adamski, N.M., Mugford, S.G., Jiang, C.-C., Febrer, M. and Uauy, C. (2012) Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biol.* **12**, 14.

Vrána, J., Kubaláková, M., Šimková, H., Číhalíková, J., Lysák, M.A. and Doležel, J. (2000) Flow sorting of mitotic chromosomes in common wheat *(Triticum aestivum* L.*). Genetics*, **156**, 2033–2041.

Wenzl, P., Suchánková, P., Carling, J., Šimková, H., Huttner, E., Kubaláková, M., Sourdille, P., Paul, E., Feuillet, C., Kilian, A. and Doležel, J. (2010) Isolated chromosomes as a new and efficient source of DArT markers for the saturation of genetic maps. *Theor. Appl. Genet.* **121**, 465–474.

Wicker, T., Taudien, S., Houben, A., Keller, B., Graner, A., Platzer, M. and Stein, N. (2009) A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. *Plant J.* **59**, 712–722.

Wicker, T., Mayer, K.F.X., Gundlach, H., Martis, M., Steuernagel, B., Scholz, U., Šimková, H., Kubaláková, M., Choulet, F., Tauden, S., Platzer, M., Feullet, C., Fahima, T., Budak, H., Dolezel, J., Keller, B. and Stein, N. (2011) Frequent gene movement and pseudogene evolution is common to the large and

complex genomes of wheat, barley and their relatives. *Plant Cell*, **23**, 1706–1718.

Supporting information

Additional Supporting information may be found in the online version of this article:

Figure S1 Sequence coverage of contigs with coding sequences.
Figure S2 Arina and Forno SNP allele distribution in 44 genotyped wheat varieties.
Table S1 Primer sequences for developed SNP and InDel markers.
Table S2 Genotyping results of Arina, Forno and 44 additional wheat varieties tested with a subset of 48 developed SNP markers.
Table S3 Ratio of detected genes with SNP to the total number of genes calculated using sliding windows.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.